

Д.К. СТУКАЛ*

**АНАЛИЗ СУБЪЕКТИВНЫХ ДАННЫХ
В ПОЛИТИЧЕСКИХ ИССЛЕДОВАНИЯХ:
ОТ ЭКСПЕРТНЫХ ОЦЕНОК
ДО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА¹**

Аннотация. Эмпирические исследования в сравнительной политологии и международных отношениях вынуждены зачастую опираться не только на собственно статистические данные, но и на экспертные оценки. Используемые при этом методы анализа данных обычно не учитывают существенные различия статистических данных и экспертных оценок, игнорируя дополнительную неопределенность, присущую последним. Данная статья посвящена обсуждению современного состояния методов сбора и обработки экспертных оценок в политологических исследованиях, а также открытых и дискуссионных вопросов в этой области.

Автор представляет байесовские процедуры анализа данных как наиболее естественный подход к обработке данных субъективной природы и акцентирует внимание на отличиях байесовского и классического подходов к анализу данных. Также рассматриваются методы получения экспертных оценок через процедуры выявления априорных распределений в целях дальнейшего использования этих распределений в байесовском анализе данных. Существующие подходы иллюстрируются примерами из проекта «Политический атлас современного мира 2.0».

* Стукал Денис Константинович, PhD, кандидат политических наук, доцент Департамента политики и управления Факультета социальных наук, Национальный исследовательский университет «Высшая школа экономики» (Москва, Россия), e-mail: dstukal@hse.ru.

¹ Статья подготовлена в рамках консорциума МГИМО МИД России и НИУ ВШЭ из средств гранта на реализацию программы стратегического академического лидерства «Приоритет-2030».

В статье обсуждаются и возможности отказа от сбора экспертных оценок в пользу «распределенного кодирования», т.е. процедур разметки качественных признаков неэкспертами на основе формализованных инструкций. В статье приводятся как успешные примеры использования «распределенного кодирования», так и сложности, стоящие на пути интеграции этого подхода в исследовательскую практику в области сравнительной политологии и международных отношений.

Наконец, завершающий раздел статьи посвящен интеграции экспертных оценок, с одной стороны, и технологий искусственного интеллекта и машинного обучения – с другой. Указывается на их совместимость в рамках байесовского подхода к анализу данных.

Ключевые слова: экспертные оценки; байесовская статистика; искусственный интеллект; методы; данные; Политический атлас современного мира 2.0.

Для цитирования: Стукал Д.К. Анализ субъективных данных в политических исследованиях: от экспертных оценок до искусственного интеллекта // Политическая наука. – 2024. – № 2. – С. 37–54. – DOI: <http://www.doi.org/10.31249/poln/2024.02.02>

Субъективность в строгости: постановка проблемы

Широкое распространение вычислительных методов в сравнительных политологических исследованиях, фиксируемое на страницах некоторых российских и международных научных журналов, с неизбежностью ставит вопрос об источниках анализируемых данных. В некоторых областях политической науки (например, в исследованиях политического поведения и коммуникации) проблема снимается за счет использования новых типов данных: например, анализа больших объемов текстовых данных из социальных сетей и мессенджеров, дающих срез наблюдаемого поведения и высказываний широкого круга пользователей (пусть и не всегда репрезентативного с точки зрения интересующей исследователя генеральной совокупности). Соответствующие исследования зачастую фокусируются на одной стране (или даже субнациональном регионе), не претендуя на широкие кросс-страновые сопоставления. Подобное сужение фокуса исследования понятно: сбор больших объемов данных по большому числу государств неизбежно порождает новые вызовы, связанные с доступом к данным, их стандартизацией, решением проблемы языка и прочее. При решении же классических задач сравнительной политологии и международных отношений, связанных с сопоставительным анализом государств, указанные выше источники данных не всегда пригодны и доступны. Объективных данных статистического характера мо-

жет быть недостаточно, в связи с чем возникает потребность в экспертных оценках, способы получения и обработки которых требуют особых подходов и инструментов. Как обрабатывать такие данные и как учитывать их субъективный характер? Можно ли с помощью методов анализа данных снизить субъективность данных или, наоборот, использовать субъективные данные для валидации результатов анализа статистических данных? Обсуждению этих вопросов посвящена статья.

Следует признать, что статистические данные о государствах мира очень разнообразны и охватывают самый широкий круг тем: от электоральных данных¹ и данных о конституционном дизайне² до макроэкономической и социальной статистики³. При этом, с одной стороны, существующие базы страновых данных не лишены целого ряда проблем и недостатков [Эмпирические вызовы..., 2023, «Политический атлас современного мира 2.0»..., 2023], а с другой – зачастую содержат в себе не только или даже не столько статистические данные, сколько экспертные оценки. Примером может быть база данных о нарушениях прав человека CIRI⁴, в которой собраны результаты ручного кодирования информации из страновых отчетов Государственного департамента США профессиональными кодировщиками, участвующими в проекте CIRI. Оставляя за рамками обсуждения вопросы о единственности и объективности используемого источника данных, обратим лишь внимание на то, что количественные данные, содержащиеся в этой базе данных, имеют субъективный характер и отражают результат прочтения текстового источника информации узким кругом людей. Тем не менее к данным проекта CIRI применяются те же самые методы анализа, что и к макроэкономической и иной статистике: переменные из базы CIRI могут, например, использоваться в качестве объясняющих признаков в регрессионных моделях – наравне с макроэкономическими и другими статистическими данными.

¹ MIT Election Data. – Mode of access: <https://electionlab.mit.edu/> (accessed: 25.01.2024).

² Comparative Constitutions Project. – Mode of access: <https://comparativeconstitutionsproject.org/> (accessed: 25.01.2024).

³ World Bank Open Data. – Mode of access: <https://data.worldbank.org/> (accessed: 25.01.2024).

⁴ Cingranelli D.L., Richards D.L., Clay K.Ch. The CIRI Human Rights Dataset. – 2014. – Mode of access: <http://www.humanrightsdata.com> (accessed: 25.01.2024).

При таком подходе, однако, игнорируется существенная неопределенность подобных данных, что сложно признать корректным подходом.

В данной статье рассматриваются три основных круга вопросов. Во втором разделе в контексте важности экспертных оценок обсуждаются байесовские методы анализа данных. Третий раздел посвящен классическим и современным подходам к получению экспертных оценок, их возможностям и ограничениям (в том числе применительно к совместному проекту МГИМО – НИУ ВШЭ «Политический атлас современного мира 2.0»). В четвертом и пятом разделах проблематика субъективных данных выводится за рамки собственно экспертных оценок путем обсуждения вопросов распределенного кодирования данных и использования методов искусственного интеллекта.

Экспертное знание в статистике: от классики до байесовских методов

Активные исследования методов сбора и анализа экспертных оценок начались в нашей стране в 1960-е годы и привели к формированию устойчивого научного сообщества вокруг регулярного научного семинара «Экспертные оценки и анализ данных», созданного под руководством П.Ф. Андруковича, Б.Г. Литвака и Ю.Н. Тюрина на механико-математическом факультете МГУ [Орлов, 2013] и существующего по сей день под руководством Ф.Т. Алескерова и Д.А. Новикова на базе Института проблем управления им. В.А. Трапезникова РАН. Если в Советском Союзе и России развитие этого направления внутри сообщества специалистов по анализу данных акцентировало внимание на методах анализа нечисловой информации [Орлов, 2013], то в США и Европе оно в значительной мере развивалось в русле параметрических методов статистики и в 2000-х годах оказалось тесно интегрированным в теорию байесовских методов анализа данных. С учетом растущей популярности байесовских методов в сравнительных политологических исследованиях, мы рассмотрим в третьем разделе, в первую очередь, связь анализа экспертных оценок с байесовскими методами.

Типичная задача анализа данных требует оценки или проверки гипотез о некоторых параметрах (характеристиках) гене-

ральной совокупности. Так, исследователя может интересовать величина эффекта государственной состоятельности (измеренной каким-либо индексом) на способность государства обеспечивать устойчивый экономический рост. В этом случае исследовательский вопрос можно формализовать в виде уравнения регрессии, где зависимая переменная – экономический рост, объясняющая переменная – индекс государственной состоятельности, а коэффициент при государственной состоятельности отражает интересующий нас эффект (мы оставляем за рамками обсуждения смежные вопросы о том, в каких случаях и при каких допущениях коэффициенты при переменных разумно интерпретировать в терминах причинных эффектов). В рамках традиционного подхода задача исследователя – максимально полно использовать всю содержащуюся в данных информацию, чтобы наиболее корректно оценить (т.е. примерно рассчитать) искомый коэффициент. Естественно, однако, что у разных государств соотношение государственной состоятельности и экономического роста будет разным. Классическая интерпретация этих различий апеллирует к идее генеральной совокупности: в среднем существует закономерная связь между уровнем государственной состоятельности и темпами экономического роста, но в каждом отдельном случае (после учета различных мешающих факторов) эта взаимосвязь подвержена некоторым случайным ошибкам. Именно наличие таких случайных ошибок и объясняет, почему коэффициенты регрессии на основе собранных данных могут быть лишь оценены, т.е. вычислены приблизительно, а не точно. Иными словами, к вычисленным на основе анализа данных числовым значениям традиционная статистика относится аналогично тому, как Платон относился к наблюдаемым объектам физического мира, т.е. как к неидеальным отражениям истинной сущности. Вследствие такого отношения возникает задача *статистического вывода* (*statistical inference*) – сделать выводы об истинном коэффициенте регрессии, существующем в предполагаемой генеральной совокупности, на основе рассчитанной по выборке оценки коэффициента. Эта задача решается либо построением доверительного интервала для коэффициента генеральной совокупности, либо проверкой статистической гипотезы (например, о том, что эффект равен нулю). В обоих вариантах ключевая информация для статистического вывода берется из вариации (дисперсии) выборочной оценки коэффициента, обусловленной описанными выше случайными ошиб-

ками. Заметим, что в рамках этого подхода вся информация извлекается из наблюдаемых данных, а исходные допущения и ожидания исследователя формально никак не учитываются (проявляясь, впрочем, в том, какие именно данные исследователь собрал и какие переменные включил в регрессионную модель). В отличие от этих, классических, процедур статистического анализа, байесовские методы дают возможность отказаться от слепой веры в данные как единственный источник информации и формализовать учет исходных представлений исследователя.

Байесовские методы анализа данных основаны на совмещении информации, содержащейся в данных, с исходной (*априорной*) информацией, имеющейся у исследователя. Такая априорная информация формально выражается в виде распределений вероятностей на множестве возможных значений коэффициента регрессии. Например, исследователь может ожидать, что экономический рост зависит не от государственной состоятельности, а от конкретной политики, проводимой Центральным банком и экономическим блоком правительства (т.е. ожидать нулевого эффекта государственной состоятельности); если же какой-то эффект государственной состоятельности все же есть, то он небольшой и, вероятно, положительный. Такие априорные представления могут быть выражены несимметричным распределением вероятностей, сконцентрированным в положительной полуплоскости, и с максимумом в нуле.

Априорные представления (выраженные в виде распределения вероятностей) совмещаются с информацией в собранных данных (формализованной через т.н. функцию правдоподобия) с помощью формулы Байеса, порождая *апостериорное* распределение. Это новое распределение вероятностей указывает, как нужно думать про наиболее вероятные значения коэффициента регрессии с учетом как собранных данных, так и исходных представлений исследователя. Упрощая логику байесовского анализа, авторы популярного учебника по байесовской статистике сформулировали «байесовскую мантру»: «апостериорное распределение пропорционально произведению априорного распределения и правдоподобия» [Gelman et al., 2020].

Апостериорное распределение вероятностей в значительной мере можно считать итоговым продуктом байесовского анализа. Но в его получении важную роль играет распределение априорное. Откуда оно берется? Здесь можно выделить два принципиально

разных подхода. В рамках первого – наиболее часто используемого – подхода априорные распределения выбираются фактически вне зависимости от реальных исходных представлений исследователя. В таком случае использование априорных распределений решает техническую задачу – предотвратить чрезмерную подгонку результатов под собранные данные (на языке машинного обучения такая чрезмерная подгонка именуется переобучением). Ряд результатов в статистической литературе указывает на тесную связь байесовских процедур с методами регуляризации, нацеленными на предотвращение переобучения и повышение обобщающей способности получаемых результатов [Hastie et al., 2016]. Таким образом, хотя априорные распределения в рамках первого подхода и не позволяют выразить и учесть в анализе исходные идеи автора, их использование может иметь важные положительные эффекты.

Наоборот, второй подход ставит своей задачей корректно отразить в априорном распределении исходные представления исследователя и учесть их при получении итогового результата – апостериорного распределения. Для решения этой задачи, однако, исходные представления исследователя необходимо для начала выявить (*prior elicitation*). Именно здесь и оказываются востребованными методы работы с экспертными оценками.

Экспертные оценки: как их получить и что с ними делать?

Традиционные методы сбора экспертных оценок, включающие в себя широкий круг процедур от различных видов интервью до мозговых штурмов и деловых игр [Экспертные оценки..., 1977], могут быть с трудом адаптированы под задачи построения априорных распределений для байесовских процедур анализа данных. В качестве альтернативы в литературе по выявлению экспертных представлений был предложен ряд подходов, основанных на сборе с экспертов информации об их неопределенности относительно измеряемой величины с последующим построением агрегированного априорного распределения [Quigley et al., 2018]. Эти подходы включают в себя три этапа: 1) измерение неопределенности экспертов, 2) восстановление априорного распределения эксперта, 3) агрегирование экспертных априорных распределений.

Вероятно, ключевое новшество обсуждаемых подходов состоит в том, что для измерения неопределенности экспертов просят указать наиболее вероятный диапазон значений, в котором лежит неизвестная оцениваемая величина. Точнее, экспертам может быть предложено указать диапазон, в котором искомая величина лежит с вероятностью 0,9. Выражаясь более строгим языком, с экспертов собирается информация о квантилях их априорного распределения: квантилях уровня 0,05 и 0,95, а также медиане. В Приложении 1 приведен фрагмент разработанной анкеты экспертного опроса для сбора данных в рамках проекта «Политический атлас современного мира 2.0». Задача опроса – выявить априорные представления о значениях индекса государственности в 10 странах мира.

К сожалению, на основе полученных от экспертов на первом этапе квантилей невозможно однозначно восстановить распределение вероятностей [Gosling, 2018]: одним и тем же квантилям может соответствовать огромное множество различных распределений. По этой причине исследователю необходимо сделать дополнительные допущения об общем виде (и конкретном семействе) распределения. Если измеряемая величина принимает значения на большом интервале (например, от 0 до 100) и мыслится симметричным, то может подойти Гауссово распределение; если оно мыслится скошенным вправо, то логнормальное; если величина принимает значения в интервале от 0 до 1, то можно использовать распределения из очень гибкого семейства бета-распределений. Этот перечень можно продолжать. В этом контексте неудивительно, что в литературе многократно указывалось, что полезно – при наличии такой возможности – просить экспертов прокомментировать, какого вида распределение они себе представляют, когда думают об измеряемой величине [Gosling et al., 2012]. После принятия решения о том, распределением какого семейства следует описывать априорные представления эксперта, точные параметры этого распределения могут быть найдены с помощью метода наименьших квадратов [O'Hagan, 1998]. Таким образом, итогом второго этапа является восстановленное априорное распределение, характеризующее мнение эксперта об измеряемой величине.

Наконец, на третьем этапе решается задача агрегирования априорных распределений индивидуальных экспертов в общее априорное распределение. Для этого обычно используются процеду-

ры объединения мнений (*opinion pooling*), среди которых наиболее распространенными являются линейная и логарифмическая процедуры [Stone, 1961]. Линейная процедура получает итоговое априорное распределение f_g как взвешенное среднее индивидуальных экспертных распределений: $f_g = \sum_j w_j f_j$, где f_j – распределение j -того эксперта, w_j – вес j -того эксперта; логарифмическая процедура использует формулу: $f_g = c \prod_j (f_j)^{w_j}$, где c – нормировочный множитель, необходимый для того, чтобы интеграл от плотности распределения равнялся единице [Rufo et al., 2012].

Ценность экспертных распределений состоит не только в том, что они могут использоваться в качестве априорного распределения в байесовских методах анализа данных (в частности, при построении индексов). Такие распределения можно также использовать, с одной стороны, для валидации результатов моделирования, а с другой – для оценки качества экспертного знания. Первая из упомянутых задач, очевидно, состоит в сравнении полученных модельных результатов (регрессионного коэффициента или рассчитанного индекса) с экспертными оценками: результаты эмпирических расчетов могут сравниваться со средним значением или максимумом априорного экспертного распределения; другая возможность – сравнение эмпирических результатов с интервалом наибольшей плотности (*highest density interval*) – байесовским аналогом доверительного интервала. Вторая задача не менее важна и может рассматриваться в качестве логического развития вопроса о согласованности экспертов – традиционной темы в области анализа экспертных оценок [Экспертные оценки..., 1977]. В современной постановке вопрос о качестве экспертных оценок включает в себя два аспекта: информативность и откалиброванность. Под информативностью (*sharpness*) понимается то, насколько эксперт готов высказывать суждения, отличные от тривиальных. Например, при ответе на вопрос (Приложение 1) неинформативным можно признать ответ эксперта, в котором нижняя и верхняя границы интервала указываются равными минимальному и максимальному значениям индекса государственности (0 и 10 соответственно). Откалиброванность же – это соответствие между внутренней шкалой, которой пользуется эксперт, и шкалой, используемой в исследовании. Простой подход к измерению откалиброванности шкалы эксперта состоит в следующем: эксперта опрашивают о величинах, значения которых уже известны и которые могут служить своего

рода якорями (*seed*) для ненаблюдаемой внутренней шкалы эксперта; далее рассматривается расхождение между частотами наступления событий, ожидаемыми экспертами (и выраженными в виде квантилей) и наблюдаемыми эмпирически – например, с помощью критерия хи-квадрат [Quigley et al., 2018]. Пример калибрующего вопроса из анкеты, разработанной в рамках проекта «Политический атлас современного мира 2.0», приведен в Приложении 2.

Вопрос откалиброванности внутренних шкал экспертов тесно связан с проблемой смещений (*biases*) в экспертном знании – сюжету, которому посвящена большая литература как опросно-экспериментального [Tetlock, 2006], так и теоретического характера [Perälä et al., 2020].

Эксперты без экспертизы: распределенное кодирование в сборе данных

Распространение онлайн-платформ, позволяющих делегировать выполнение определенных рутинных операций произвольно широкому кругу лиц, снабженных инструкциями и примерами, привело к росту популярности подхода, получившего название краудсорсинга в науке [Lenart-Gansinieć et al., 2023]. Применительно к задаче сбора данных можно говорить о «распределенном кодировании» (*crowdcoding*), в рамках которого чаще всего качественные признаки размечаются группами неэкспертов. Такая процедура получила распространение в исследовательских проектах, ориентированных на формирование кросс-страновых данных; например, при кодировании партийных программ [Benoit et al., 2016; Lehmann, Zobel, 2018]. Приведенный пример особенно показателен, поскольку задача кодирования партийных программ традиционно рассматривалась в качестве задачи, требующей экспертного знания, и опиралась либо на экспертные опросы [Laver, Budge, 1992], либо на использование массивов данных, собранных узкими исследовательскими коллективами, включающими экспертов и их ассистентов [Franzmann, Kaiser, 2006].

Привлекательность распределенного кодирования можно объяснить несколькими факторами. Во-первых, оно избавляет от необходимости поиска узкоспециализированных экспертов, готовых принять участие в исследовании. Во-вторых, отказ от экспертного

знания в целом может снизить остроту проблемы экспертных смещений. Наконец, распределенное кодирование в силу своей природы с неизбежностью требует разработки подробных, понятных и воспроизводимых инструкций, руководствуясь которыми разметчики смогут закодировать необходимые данные. Само по себе создание таких инструкций можно считать золотым стандартом разметки, позволяющим исследователям в ясном и однозначном виде сформулировать определения понятий и процедуры их измерения.

Поскольку распределенное кодирование сводится к рутинизированному выполнению заранее сформулированных инструкций, успешность выполнения задачи в этом случае может быть количественно измерена с помощью различных метрик согласованности. Наибольшую популярность (в том числе, благодаря их доступности в стандартных библиотеках анализа данных) получили альфа Криппендорфа [Krippendorff, 1980] и каппы Фляйса и Коэна [Fleiss et al., 1969]. Тем не менее даже высокая согласованность разметчиков необязательно означает содержательной валидности измерения: сформулированные для разметчиков инструкции могут содержать в себе экспертные заблуждения и смещения, которые будут воспроизводиться в рамках распределенного кодирования. Таким образом, несмотря на изначальный оптимизм сторонников распределенного кодирования [Benoit et al., 2016], его сложно признать панацеей от всех проблем экспертного знания.

Эксперты и / или искусственный интеллект

Распространение технологий искусственного интеллекта и алгоритмов машинного обучения в социальных науках [Athey, Imbens, 2019; Molina, Garip, 2019; Edelmann et al., 2020; Grimmer et al., 2021; Brand et al., 2023] поставило новые вопросы перед практикой применения экспертного знания и распределенного кодирования. Дискуссия развернулась о сравнительных преимуществах использования (не)экспертных оценок, с одной стороны, и предсказаний алгоритмов машинного обучения – с другой [Dressel, Farid, 2018; Bansak, 2019]. Сторонники распределенного кодирования указывают на то, что эти процедуры способны достигать точности, сопоставимой с алгоритмической, в связи с чем под вопрос ставится необходимость применения предиктивных моделей в социальных науках [Dressel, Farid,

2018]. Их оппоненты, анализируя те же данные, указывают на преимущества предиктивных моделей с точки зрения откалиброванности предсказаний [Bansak, 2019]. Существенно более радикальный взгляд состоит в том, что использование технологий искусственного интеллекта желательно не только в политологических исследованиях, но и при принятии политических решений, поскольку такие технологии «имеют потенциал повысить политическую легитимность, выявляя острые общественные вопросы, предсказывая возможные последствия проводимой политики и оценивая эффективность политики» [Starke, Luenich, 2020]. Этот тезис актуализирует получившие популярность в 1980-е годы исследования в сфере разработки систем поддержки принятия решений [Ларичев, Петровский, 1987], однако на новой программной и информационной основе.

Современное развитие байесовских методов машинного обучения, интегрирующих процедуры построения предиктивных моделей с байесовским подходом к анализу данных, указывает, однако, на то, что экспертные оценки и технологии искусственного интеллекта не только не противоречат друг другу, но и могут продуктивно совмещаться. В сравнительных политологических исследованиях байесовские процедуры снижения размерности могут использоваться для построения интегральных индексов. Предсказания предиктивных моделей, построенных на основе эмпирических данных, могут сравниваться с экспертными оценками в логике апостериорной предиктивной оценки моделей [Gelman et al., 1996]. Совершенствование процедур получения обобщенных экспертных оценок, таким образом, не теряет своей актуальности и в условиях распространения искусственного интеллекта.

Заключение: строгость и субъективность в анализе данных

В ставшей уже классикой статье о двух культурах статистического моделирования Л. Брейман описал свой взгляд на хорошие практики анализа данных: сначала «вжиться» в данные и лишь затем приступить к их анализу, фокусироваться на нахождении «хороших» решений, добиваться точности модельных предсказаний [Breiman, 2001]. Большую роль в этой практике, очевидно, играет субъективность: она проявляется и во «вживании» в данные, и в

определении того, где проходит граница между «хорошими» и «плохими» статистическими результатами.

В еще более острой форме на роль субъективности в анализе данных указал Э. Лимер: «Искусство эконометрики – в том виде, в котором оно реализуется на компьютере – состоит в построении многих, возможно, тысяч статистических моделей. Одна или несколько из них, которые понравятся исследователю, будут выбраны для публичного представления» [Leamer, 1983, p. 36]. Означает ли это, что эмпирические исследования в политической науке неизбежно столь же субъективны, как и оторванные от эмпирики рассуждения о политике? А процедуры и результаты анализа *hard data* по своей сути не отличаются от анализа (не)экспертных оценок?

Сложившаяся в последние десятилетия практика проверки устойчивости результатов анализа (*robustness checks*), применение методов проверки чувствительности результатов к нарушению допущений (*sensitivity analysis*), о важности которых писал процитированный выше Э. Лимер и которые продолжают развиваться в эмпирической политологии и эконометрике [Imai, Yamamoto, 2013; Oster, 2019], позволяют дать отрицательные ответы на поставленные вопросы. Корректное и ответственное применение процедур анализа данных позволяет существенно снизить пространство субъективного. Однако безответственное применение анализа данных на это неспособно.

Строгие процедуры обработки субъективных экспертных и неэкспертных оценок, а также опирающееся на теорию байесовского анализа включение исходных (априорных) представлений в процесс обработки данных позволяют иначе взглянуть на вопрос субъективности в анализе эмпирики: субъективное знание, выявленное обсуждавшимися выше процедурами, соответствующим образом формализованное и включенное в процесс анализа данных, – это не помеха, а важный компонент строгого подхода к работе с эмпирикой, допускающего итерационное взаимодействие человека, данных и искусственного интеллекта.

D.K. Stukal*

**Subjective data in political science research:
from expert evaluation to artificial intelligence¹**

Abstract. Empirical research in Comparative Politics and International Relations is often built not only on statistical data, but also on expert evaluation data. However, the methods of data analysis employed in this case often fail to account for the differences between statistical and expert evaluation data, and disregard the extra uncertainty in the latter. This article focuses the state-of-the-art methods for collecting and processing expert evaluation data in political science research, as well as open questions in this area. The article presents Bayesian data analysis as the most natural approach to analyzing subjective data and focuses on the differences between Bayesian and classical approaches. Then the article focuses on the methods for obtaining expert evaluations through prior elicitation for further use in Bayesian analysis. These approaches are illustrated using examples from the research project “Political Atlas of the Modern World 2.0”. The next section discusses the possibility of replacing expert evaluation data with crowdcoding, i.e. the procedures for annotating or coding qualitative features by non-experts based on formalized instructions. The article cites both successful examples of crowdcoding usage in empirical research and potential challenges for its integration into research in Comparative Politics and International Relations. Finally, the author addresses the issues of integrating expert evaluation data, on the one hand, and artificial intelligence and machine learning technologies, on the other. We highlight their compatibility in the framework of Bayesian data analysis.

Keywords: expert evaluation; Bayesian statistics; artificial intelligence; methods; data; Political Atlas of the Modern World 2.0.

For citation: Stukal D.K. Subjective data in political science research: from expert evaluation to artificial intelligence. *Political science (RU)*. 2024, N 2, P. 37–54. DOI: <http://www.doi.org/10.31249/poln/2024.02.02>

References

- Athey S., Imbens G.W. Machine learning methods that economists should know about. *Annual review of economics*. 2019, Vol. 11, P. 685–725. DOI: <https://doi.org/10.1146/annurev-economics-080217-053433>
- Bansak K. Can nonexperts really emulate statistical learning methods? A comment on “The accuracy, fairness, and limits of predicting recidivism.” *Political analysis*. 2019, Vol. 27, N 3, P. 370–380. DOI: <https://doi.org/10.1017/pan.2018.55>

* **Stukal Denis**, HSE University (Moscow, Russia), e-mail: dstukal@hse.ru

¹ The article was prepared within the consortium of MGIMO University and HSE University and funded by the grant for the implementation of the Priority 2030 Strategic Academic Leadership Program.

- Benoit K., Conway D., Lauderdale B.E., Laver M., Mikhaylov S. Crowd-sourced text analysis: reproducible and agile production of political data. *American political science review*. 2016, Vol. 110, N 2, P. 278–295. DOI: <https://doi.org/10.1017/s0003055416000058>
- Brand J.E., Zhou X., Xie Y. Recent developments in causal inference and machine learning. *Annual Review of Sociology*. 2023, Vol. 49, P. 81–110. DOI: <https://doi.org/10.1146/annurev-soc-030420-015345>
- Breiman L. Statistical modeling: the two cultures. *Statistical Science*. 2001, Vol. 16, N 3, P. 199–215. DOI: <https://doi.org/10.1214/ss/1009213726>
- Dressel J., Farid H. The accuracy, fairness, and limits of predicting recidivism. *Science advances*. 2018, Vol. 4, N 1, P. eaao5580. DOI: <https://doi.org/10.1126/sciadv.aao5580>
- Edelmann A., Wolff T., Montagne D., Bail C.A. Computational social science and sociology. *Annual review of sociology*. 2020, Vol. 46, P. 61–81. DOI: <https://doi.org/10.1146/annurev-soc-121919-054621>
- Fleiss J.L., Cohen J., Everitt B.S. Large sample standard errors of kappa and weighted kappa. *Psychological bulletin*. 1969, Vol. 72, N 5, P. 323–327. DOI: 10.1037/h0028106
- Franzmann S., Kaiser A. Locating political parties in policy space: a reanalysis of Party Manifesto data. *Party politics*. 2006, Vol. 12, N 2, P. 163–188. DOI: <https://doi.org/10.1177/1354068806061336>
- Gelman A., Carlin J.B., Stern H.S., Dunson D.B., Vehtari A., Rubin D.B. *Bayesian data analysis*. New York, Chapman and Hall/CRC, 2020, 675 p.
- Gelman A., Meng X.-L., Stern H. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*. 1996, Vol. 6, P. 733–807.
- Gosling J.P. SHELF: The Sheffield Elicitation Framework. In: Dias L.C., Morton A., Quigley J. (eds). *Elicitation: the science and art of structuring judgement*. New York: Springer, 2018, P. 61–93.
- Gosling J.P., Hart A., Mouat D., Sabirovic M., Scanlon S., Simmons A. Quantifying experts' uncertainty about the future cost of exotic diseases. *Risk analysis*. 2012, Vol. 32, N 5, P. 881–893. DOI: <https://doi.org/10.1111/j.1539-6924.2011.01704.x>
- Grimmer J., Roberts M.E., Stewart B.M. Machine learning for social science: an agnostic approach. *Annual review of political science*. 2021, Vol. 24, P. 395–419. DOI: <https://doi.org/10.1146/annurev-polisci-053119-015921>
- Hastie T., Tibshirani R., Friedman J. *The elements of statistical learning: data mining, inference, and prediction*. New York: Springer, 2016, 745 p.
- Imai K., Yamamoto T. Identification and sensitivity analysis for multiple causal mechanisms: revisiting evidence from framing experiments. *Political analysis*. 2013, Vol. 21, N 2, P. 141–171. DOI: <https://doi.org/10.1093/pan/mps040>
- Krippendorff K. *Content analysis: an introduction to its methodology*. Beverly Hills, CA: Sage, 1980, 441 p.
- Larichev O.I., Petrovskiy A.B. Decision support systems: state of the art and perspectives for development. *Results of science and technology*. Moscow: VINITI, 1987, Vol. 21, P. 131–164. (In Russ.)
- Laver M., Budge I. *Party policy and government coalitions*. New York: St. Martins Press, 1992, 448 p.

- Leamer E.E. Let's take the con out of econometrics. *The American economic review*. 1983, Vol. 73, N 1, P. 31–43.
- Lehmann P., Zobel M. Positions and saliency of immigration in party manifestos: a novel dataset using crowd coding. *European journal of political research*. 2018, Vol. 7, N 4, P. 1056–1083. DOI: <https://doi.org/10.1111/1475-6765.12266>
- Lenart-Gansiniec R., Czakon W., Sulkowski L., Jasna Pocek J. Understanding crowdsourcing in science. *Review of managerial science*. 2023, Vol. 17, P. 2797–2830. DOI: <https://doi.org/10.1007/s11846-022-00602-z>
- Melville A.Yu., Malgin A.V., Mironyuk M.G., Stukal D.K. “Political atlas of the modern world 2.0”: formulation of the research problem. *Polis. Political studies*. 2023, N 2, P. 72–87. DOI: <https://doi.org/10.17976/jpps/2023.02.06> (In Russ.)
- Melville A.Yu., Malgin A.V., Mironyuk M.G., Stukal D.K. Empirical challenges and methodological approaches in comparative politics (through the lens of the Political Atlas of the Modern World 2.0). *Polis. Political studies*. 2023, N 5, P. 153–171. DOI: [10.17976/jpps/2023.05.10](https://doi.org/10.17976/jpps/2023.05.10) (In Russ.)
- Molina M., Garip F. Machine learning for sociology. *Annual review of sociology*. 2019, Vol. 45, P. 27–45. DOI: <https://doi.org/10.1146/annurev-soc-073117-041106>
- O'Hagan A. Eliciting expert beliefs in substantial practical applications. *Journal of the royal statistical society. Series D (The Statistician)*. 1998, Vol. 47, N 1, P. 21–35. DOI: <https://doi.org/10.1111/1467-9884.00114>
- Orlov A.I. Expert evaluation theory in our country. *Polythematic online scientific journal of Kuban State Agrarian University*. 2013, N 93, P. 1–11. (In Russ.)
- Oster E. Unobservable selection and coefficient stability: theory and evidence. *Journal of business & Economic statistics*. 2019, Vol. 37, N 2, P. 187–204. DOI: <https://doi.org/10.1080/07350015.2016.1227711>
- Quigley J., Colson A., Aspinall W., Cooke R.M. Elicitation in the classical model. In: Dias L.C., Morton A., Quigley J. (eds). *Elicitation: the science and art of structuring judgement*. Springer, 2018, P. 15–36.
- Perälä T., Vanhatalo J., Chrysafi A. Calibrating expert assessments using hierarchical gaussian process models. *Bayesian analysis*. 2020, Vol. 15, N 4, P. 1251–1280. DOI: <https://doi.org/10.1214/19-ba1180>
- Rufo M.J., Pérez C.J., Martín J. A Bayesian approach to aggregate experts' initial information. *Electronic journal of statistics*. 2012, Vol. 6, P. 2362–2382. DOI: <https://doi.org/10.1214/12-ejs752>
- Schmerling D.S., Dubrovskiy S.A., Arzhanova T.D., Frenkel A.A. Expert evaluation. Methods and application. In: *Statistical methods of expert evaluation analysis*. Moscow: Nauka, 1977, P. 290–382. (In Russ.)
- Starke C., Lünich M. Artificial intelligence for political decision-making in the European Union: effects on citizens' perceptions of input, throughput, and output legitimacy. *Data & Policy*. 2020, Vol. 2, E16. DOI: <https://doi.org/10.1017/dap.2020.19>
- Stone M. The opinion pool. *Annals of mathematical statistics*. 1961, Vol. 32, N 4, P. 1339–1342. DOI: <https://doi.org/10.1214/aoms/1177704873>
- Tetlock Ph. E. *Expert political judgment: how good is it? how can we know?* Princeton: Princeton university press, 2006, 321 p.

Литература на русском языке

- Ларичев О.И., Петровский А.Б. Системы поддержки принятия решений: современное состояние и перспективы развития // Итоги науки и техники. – М.: ВИНТИ, 1987. – Т. 21. – С. 131–164.
- Орлов А.И. Теория экспертных оценок в нашей стране // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2013. – № 93. – С. 1–11.
- «Политический атлас современного мира 2.0»: к постановке исследовательской задачи / А.Ю. Мельвиль, А.В. Мальгин, М.Г. Миронюк, Д.К. Стукал // Полис. Политические исследования. – 2023. – № 2. – С. 72–87. – DOI: <https://doi.org/10.17976/jpps/2023.02.06>
- Экспертные оценки. Методы и применение / Д.С. Шмерлинг, С.А. Дубровский, Т.Д. Аржанова, А.А. Френкель // Статистические методы анализа экспертных оценок. – М.: Наука, 1977. – С. 290–382.
- Эмпирические вызовы и методологические подходы в сравнительной политологии (сквозь призму «Политического атласа современного мира 2.0») / А.Ю. Мельвиль, А.В. Мальгин, М.Г. Миронюк, Д.К. Стукал // Полис. Политические исследования. – 2023. – № 5. – С. 153–171. – DOI: <https://doi.org/10.17976/jpps/2023.05.10>

ПРИЛОЖЕНИЯ

Приложение 1

Пример экспертного опроса об уровне государственности

Пояснение. Экспертную оценку часто проще дать не в виде одного-единственного числа, а в виде некоторого интервала вероятных значений. Ниже мы попросим Вас оценить значения различных индексов для некоторых государств, используя 90%-ные интервалы, т.е. такие интервалы, которые содержат корректное значение индекса для данного государства с вероятностью 90%.

Для таких 90%-ных интервалов мы будем просить Вас указать нижнюю и верхнюю границы, отражающие разброс возможных значений индекса для данного государства. Кроме того, мы будем просить Вас указать медиану Вашей оценки: с вероятностью 50% корректное значение оцениваемой характеристики государства будет либо меньше, либо больше этой медианы.

Вопрос. Для измерения *государственности* разрабатывается индекс, принимающий непрерывные значения от 0 до 10. Оцените, пожалуйста, уровень развития государственности у приведенных

ниже государств по состоянию на 2020 г. в непрерывной шкале от 0 до 10. Приведите, пожалуйста, медиану и границы 90%-ного интервала для Ваших оценок:

Страна	Нижняя граница 90%-ного интервала	Верхняя граница 90%-ного интервала	Медиана
Аргентина			
Бразилия			
Индия			
Индонезия			
Китай			
Россия			
США			
Турция			
Чили			
Южноафриканская республика			

Приложение 2

Пример калибрующего вопроса об уровне демократии

Вопрос. Оцените, пожалуйста, *уровень демократии в понимании индекса «Полития» (Polity V)* в приведенных ниже государствах по состоянию на 2020 г. в непрерывной шкале от –10 до +10. Приведите, пожалуйста, медиану и границы 90%-ного интервала для Ваших оценок:

Страна	Нижняя граница 90%-ного интервала	Верхняя граница 90%-ного интервала	Медиана
Аргентина			
Бразилия			
Индия			
Индонезия			
Китай			
Россия			
США			
Турция			
Чили			
Южноафриканская республика			